



Bioinformatics and its applications

Alla L Lapidus, Ph.D. SPbAU, SPbSU, St. Petersburg





Term Bioinformatics

Term **Bioinformatics** was invented by Paulien Hogeweg (Полина Хогевег) and Ben Hesper in 1970 as "the study of informatic processes in biotic systems".

Paulien Hogeweg is a Dutch theoretical biologist and complex systems researcher studying biological systems as dynamic information processing systems at many interconnected levels.

Definitions of what is Bioinformatics:

Bioinformatics is the use of IT in biotechnology for the data storage, data warehousing and analyzing the DNA sequences. In Bioinfomatics knowledge of many branches are required like biology, mathematics, computer science, laws of physics & chemistry, and of course sound knowledge of IT to analyze biotech data. Bioinform Bioinformatics is an interdisciplinary field that develops and improves upon used to solve n methods for storing, retrieving, organizing and analyzing biological data. A

major activity in bioinformatics is to develop software tools to generate useful

The mathematical, statistical and computing methods that aim to solve biological problems using DNA and amino acid

sequences and related information of computational tools

Bioinformatics develop of computer to analy biology for example ingredients and metable or application of computational tools and approaches for expanding the use of biological, medical, behavioral or health data, including those to acquire, store, organize, archive, analyze, or visualize such data.

http://www.bisti.nih.gov/CompuBioDef.pdf

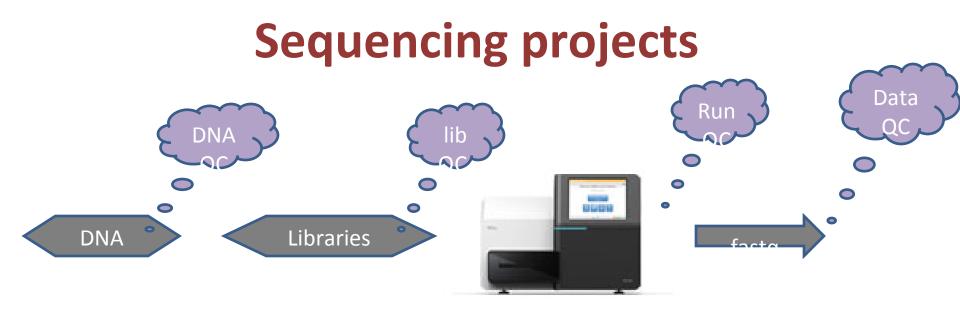
My additions:

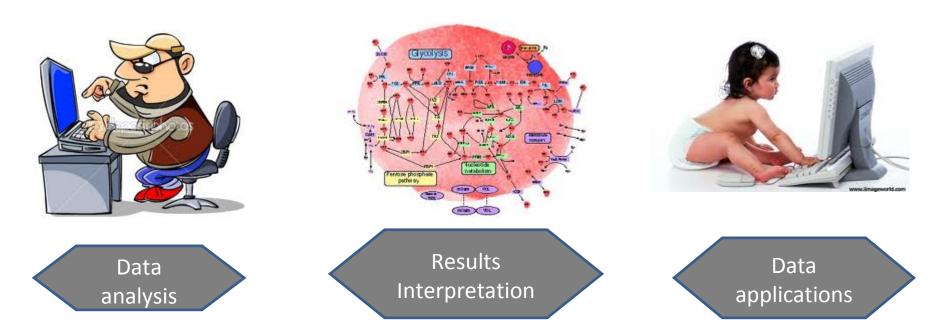
- **1. Bioinformatics is a SCIENCE**
- **2. Not only to develop algorithms, store, retrieve, organize and analyze biological data but to CURATE data**

Bioinformatics is being used in following fields:

- Microbial genome applications
- Molecular medicine
- Personalised medicine
- Preventative medicine
- Gene therapy
- Drug development
- Antibiotic resistance
- Evolutionary studies
- ✤ Waste cleanup
- Biotechnology

- Climate change Studies
- Alternative energy sources
- Crop improvement
- ✤ Forensic analysis
- ✤ Bio-weapon creation
- ✤ Insect resistance
- Improve nutritional quality
- Development of Drought resistant varieties
- Vetinary Science





LIMS - Lab Information Management Software

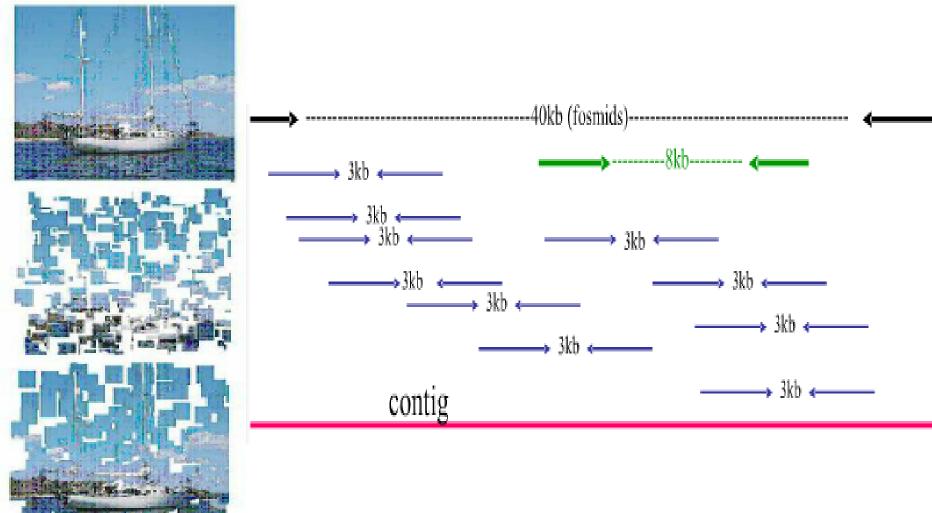
Microbial genome applications

- Genome assembly
- Re-sequencing
- Comparative analysis
- Evolutionary studies
- Antibiotic resistance
- Biotechnology

Genome Assembly

- Genome assembly is a very complex computational problem due to enormous amount of data to put together and some other reasons reasons.
- Ideally an assembly program should produce one contig for every chromosome of the genome being sequenced. But because of the complex nature of the genomes, the ideal conditions just never possible, thus leading to gaps in the genome.

De Novo assembly - puzzle without the picture

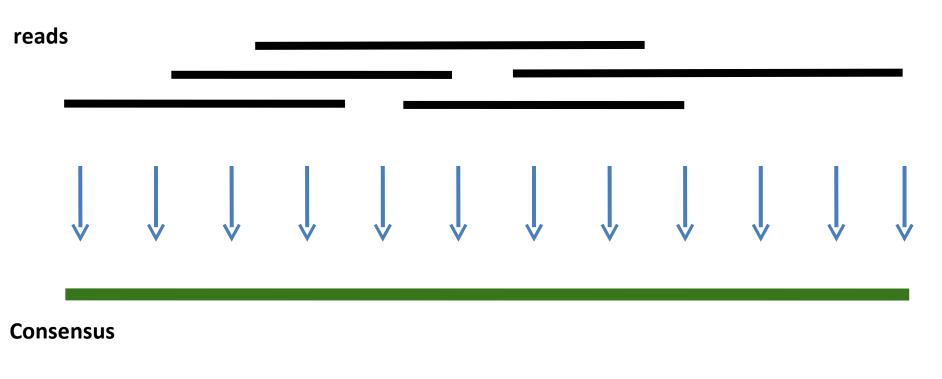


Assembly Challenges

- Presence of repeats. Repeats are identical sequences that occur in the genome in different locations and are often seen in varying lengths and in the multiple copies. There are several types of repeats: tandem repeats or interspersed repeats. The read's originating from different copies of the repeat appear identical to the assembler, causing errors in the assembly.
- **Contaminants** in samples (eg. from Bacteria or Human).
- PCR artefacts (eg. Chimeras and Mutations)
- Sequencing errors, such as "Homopolymer" errors when eg. 2+ run of same base.
- MID's (multiplex indexes), primers/adapters still in the raw reads.
- polyploid genomes

Assembly algorithms

Overlap-Layout-Consensus - Find overlaps between all reads



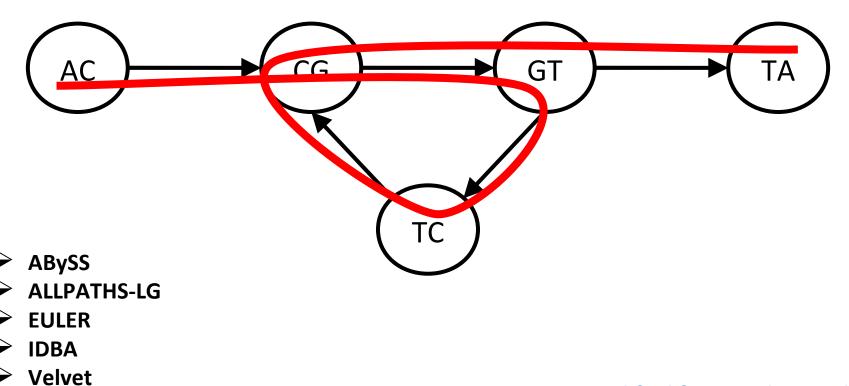
Problems caused by new sequencing technologies:

- Hard to find overlaps between short reads
- Impossible to scale up

De Bruijn graph

ACGTCGTA





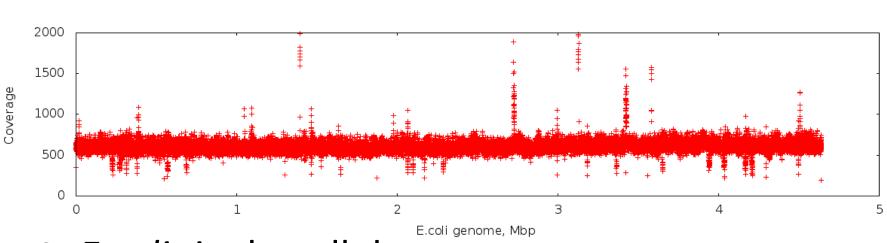
Modified from Andrey Prjibelski

Single-cell dataset

IDBA-UD

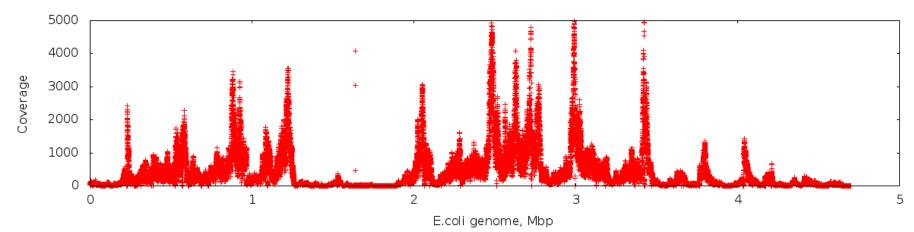
SPAdes

Velvet



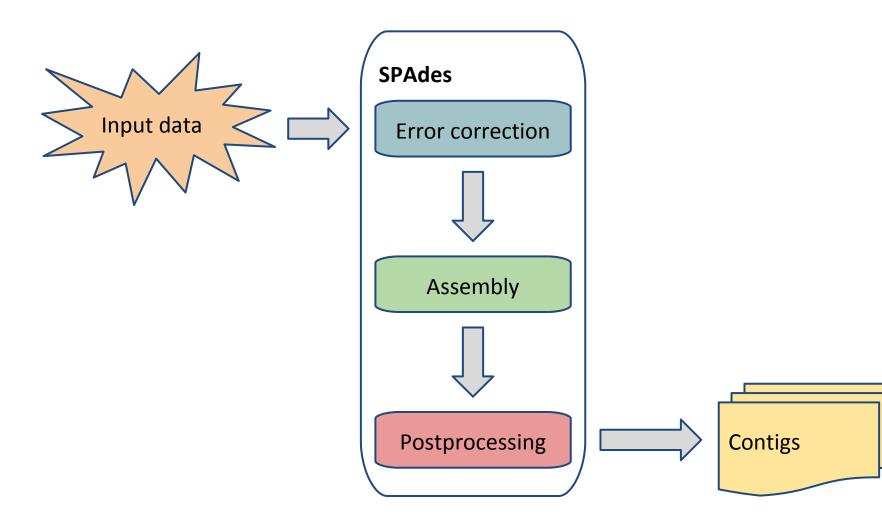
• E.coli single-cell dataset

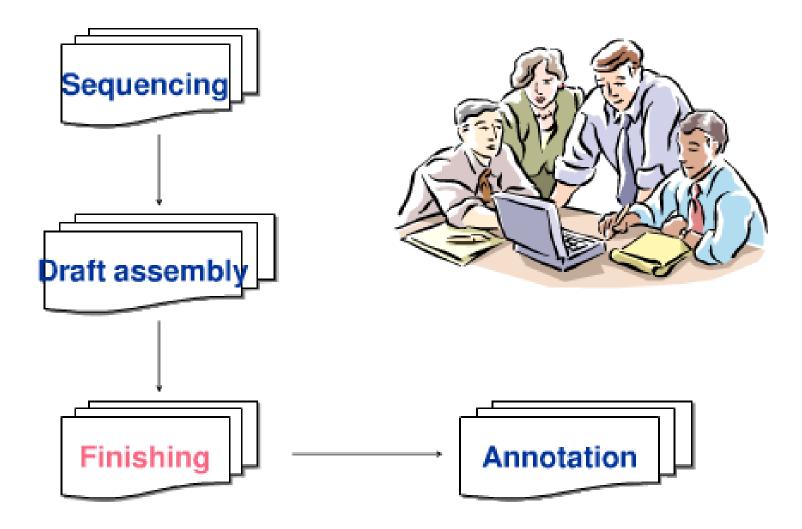
E. coli isolate dataset



SPAdes pipeline







Gene Prediction and Genome Annotation

1	aacagggtgt	atctcgcaca	ttctcatcca	ctagtataac	tgctgctgac	agtaatcgaa
61	ctagatagac	tgttctggat	gctatcattc	gatattttga	caacacggga	gccatcctgt
121	tcgttgatcc	gagattcgac	gagtcatgca	acaagatcca	gaccgttgcc	tgcaaacgcc
181	taggctgtga	atgaacgact	cgatcacgat	cgctagtcgc	acgtctgatc	tcaccgattg
241	aagcogtatt	ccacagagtg	cgagaaccgg	tcatttactg	agtggttcgg	ctctgtttaa
301	atacggaaag	cccactcggg	agagatatct	ctccttaatg	ggctatgaaa	ggtatgaatg
361	gtggcggcga	accgcgtttc	ccagaggctc	gcgcactcca	gtactccccg	gaacgctggt
421	gggcttatct	tccgtgttcg	ggatgggtac	gggaggcaac	cccaccgctg	tagcegreta
481	acgtcgagtc	acggaatcga	accgcgatag	taccagtctc	gattaactct	tccaccgigt
541	gattacgtgc	gatccagttt	gcgcctggac	tcgttcagcg	acgagttaaa	tcg <mark>atggt</mark> ga
601	atgagtcaca	gtgcgtatga	atgatggett	tggtctgtta	gtgctcgtgg	gettaacete
661	tcgttacctc	gacgcgcaca	ccccgagtct	atcgaccgcg	tcttgtacgc	gggacctcgg
721	cggtgtctct	tttccaagtg	ggtttcgagc	ttagatgcgt	tcagetetta	ccccgtgtgg
781	cgtggctacc	cggcacgtgc	tetetegaae	aaccggtaca	ccagtggcca	ccaaccgtag
841	tteetetegt	actatacggt	cgttcttgtc	agacaccatt	acacacccag	tagatagcag
901	ccgacctgtc	tcacgacggt	ctaaacccag	ctcacgacat	cctttaatag	gcgaacaacc
961	tcacccttgc	ccgcttctgc	acgggcagga	tggagggaac	cgacatcgag	gtagcaagcc
1021	actcggtcga	tatgtgctct	tgcgagtgac	gaccototta	tccctagggt	agettttetg
1081	tcatcaattg	cccgcatcaa	gcaggctaat	tggttcgdta	gaccacgett	tegegteage
1141	gtteetegtt	gggaagaaca	ctgtcaagct	taattttget	cttgcactct	tcgccgggtc
1201	tctgtcccgg	ctgagatagc	catagggcgc	gctcgatatc	ttttcgagcg	cgtaccgccc
1261	cagtcaaact	gcccggctat	cggtgtcctc	cteeeggagt	gagagtcgca	gtcaccgacg
1321	ggtagtattt	cactgttgac	tcggtggccc	gctagcgcgg	gtacctgtgt	agtgtctcct
1381	atgtatgctg	cacateggeg	accacgtete	agegacagee	tgcagtaaag	ctccataggg
1441	tettegette	cccctgggtg	tctccagact	ccgcactgga	atgtacagtt	caccgggccc
1501	aacgttggga	cagtgaaget	ctggttaatc	cattcatgca	agccgctact	gatgcggcaa
1561	ggtactacgc	taccttaaga	gggtcatagt	tacccccgcc	gttgacaggt	cettegteet
1621	cttgtacgag	gtgttcagat	acctgcactg	ggcaggattc	agtgaccgta	cgagtccttg
1681	cggatttgcg	gtcacctatg	ttgttactag	acagtccgag	cttccgagtc	actgcgacct
1741	gctccgttcc	ggagcaggca	tecettette	cgaaggtacg	ggactaactt	gccgaattcc
1801	ctaacgttgg	ttgctcccga	caggeettgg	ctttcgccgc	catggacacc	tgtgtcggtt

<u>Based on similarity to</u> <u>known genes</u> – blastX (NCBI)

Gene finding programs

- **Glimmer** for most procaryotic genomes
- GenMark for both procaryotic genomes and eucaryotic genomes

img 🕅 🕂

INTEGRATED MICROBIAL GENOMES

My IMG

IMG Home Find Genomes

Find Genes | Find Functions

Compare Genomes

omes Analysis Cart

Companion Systems

Using IMG

IMG Content Datasets 6120 Bacteria 248 Archaea 183 Eukarya 1193 Plasmids 2809 Viruses Genome Fragments 579 Total Datasets 11132 GEBA 245 Last updated: 2013-07-05

IMG 4.0 is dedicated to the memory of our colleaque, lain Anderson

Genome by Metadata Project Map Content History System Requirements About IMG FAQ

> Hands on training available at the

Microbial Genomics & Metagenomics Workshop The Integrated Microbial Genomes (IMG) system (<u>Nucleic Acids Research</u>, Vol 40, 2012) serves as a community resource for comparative analysis and annotation of all publicly available genomes from three domains of life in a uniquely integrated context. Plasmids that are not part of a specific microbial genome sequencing project and phage genomes are also included into IMG in order to increase its genomic context for comparative analysis.

For details, see <u>IMG Release Notes</u> (Dec. 12, 2012), in particular the workspace and background computation capabilities available to IMG registered users.

Count	Total
DNA, number of bases	60,476,662,654
Total Genes	25,395,838
Total Genomes	<u>11,132</u>



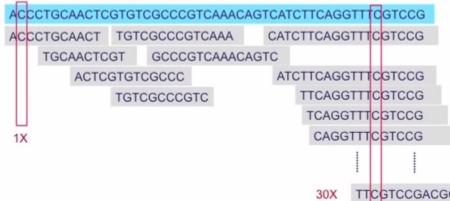
Genome Count

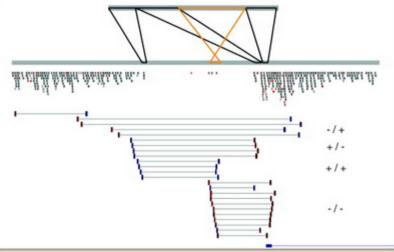
Status	Bacteria	Archaea	Eukaryota	Plasmids	Viruses	Genome Fragments	Total
Finished	<u>2131</u>	<u>154</u>	<u>37</u>	<u>1190</u>	<u>2809</u>	<u>579</u>	<u>6900</u>
Draft	<u>2407</u>	<u>28</u>	<u>146</u>	<u>3</u>	0	0	<u>2584</u>
Permanent Draft	<u>1582</u>	<u>66</u>	0	0	0	0	<u>1648</u>
Total	<u>6120</u>	<u>248</u>	<u>183</u>	<u>1193</u>	<u>2809</u>	<u>579</u>	<u>11132</u>

Re-sequencing

Projects aimed at characterizing the genetic variations of species or populations

- Resequencing of bacterial and archaeal isolates etc is possible if reference genomes are available
- This approach can help to better understand bacterial community structure, gene function in bacteria under selective pr





Climate change Studies

Increasing levels of carbon dioxide emission are thought to contribute to global climate change.

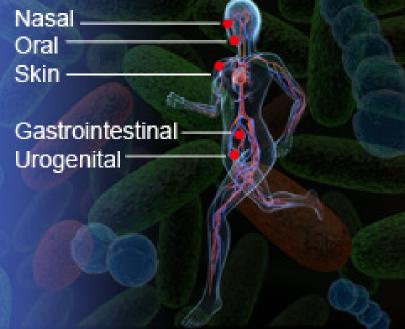
One way to decrease atmospheric carbon dioxide is to study the genomes of microbes that use carbone dioxidet as their sole carbon source

Human microbiome

MetaHIT - Europe Human Microbiome Project – US

The human microbiome includes viruses, fungi and bacteria, their genes and their environmental interactions, and is known to influence human physiology.

There's very broad variation in these bacteria in different people and that severely limits our ability to create a "normal" microflora profile for comparison among healthy people and those with any kind of health issues.



Children with **autism** harbor significantly fewer types of gut bacteria than those who are not affected by the disorder, researchers have found.

Prevotella species were most dramatically reduced among samples from autistic children—especially *P. copri*. (helps the breakdown of protein and carbohydrate foods)

Bioinformatics combining biology with computer science

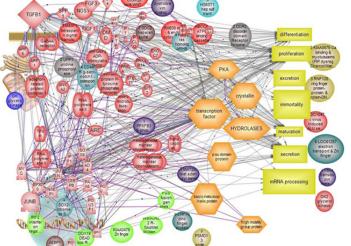
- it can explore the causes of diseases at the molecular level
- explain the phenomena of the diseases on the gene/pathway level
- make use of computer techniques (data mining, machine learning etc), to analyze and interpret data faster
- to enhance the accuracy of the results



Reduce the cost and time of drug discovery

To improve drug discovery we need to discover (read "develop") efficient bioinformatics algorithms and approaches for

> target identification target validation lead identification lead optimization

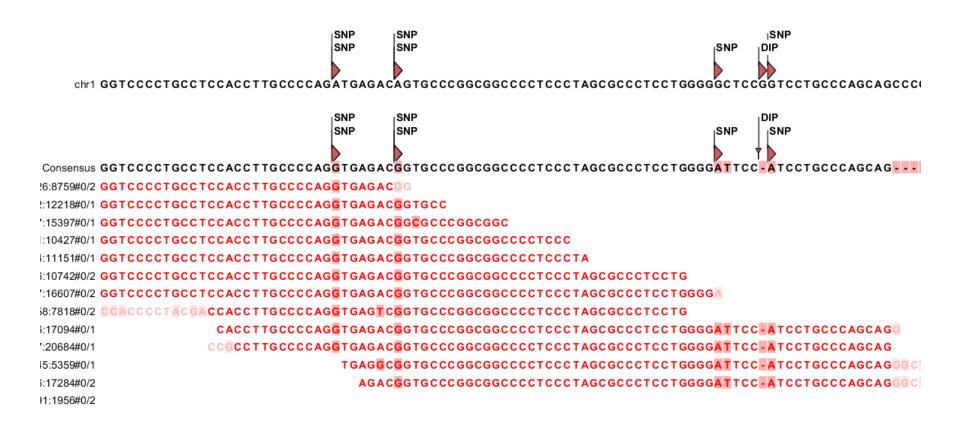


Advantages of detecting mutations with next-generation sequencing

High throughput

- Test many genes at once
- Systematic, unbiased mutation detection
 - All mutation types
 - Single nucleotide variants (SNV), copy number variation (CNV)-insertions, deletions and translocations
- Digital readout of mutation frequency
 - Easier to detect and quantify mutations in a heterogeneous sample
- Cost effective precision medicine
 - "Right drug at right dose to the right patient at the right time"

Homozygous SNPs and indel



Poor alignment

۰ III ا	
耐 🔿 🖾 🗉 🏛 🏶 🎼 🕼 📝	
🔁 [chr 1 mapping] 📀	
chr1	TGGAATGGTATGGAATGGAATGGAATGGAATGGAATGG
	SNP
Consensus 2999	TGGAATGGTATGGAATGGAATGGAATGGAATGGAATGG
;1521_0001_FC:2:60:2158:13703#0/2	TG
S1521_0001_FC:2:50:6708:6740#0/1	TGGACTAGAA
S1521_0001_FC:2:80:5495:2498#0/1	TGGAATGGTATGGAATGGAATGGAATGGAATGGAATGCCATCAACCCGA
S1521_0001_FC:2:19:8383:7452#0/1	TGGAATGGTATGGAATGGAATGGAATG
S1521_0001_FC:2:68:3943:8014#0/2	TGGAATGCTATGGAATGGAATGGAATGGAATGGAGTGCATTGGAA
S1521_0001_FC:2:82:5095:3807#0/1	TGGAATGGTATGGAATGGAATGGAATGATATGGAAGAATGAAATTGATTGAACTGGAGTG
S1521_0001_FC:2:42:7841:5804#0/1	TGGAA BGG ATGGA TGGAACAAAA GGAA BAAATGGAC BGGAATGGAATGGAATGGACTCGAATGGAATGGA
;1521_0001_FC:2:83:11412:4622#0/1	CEAATGEAATGEAATAEAETGEAATEGAATEGAATGEAATGEAATGEAATGEAATG
;1521_0001_FC:2:16:11936:9501#0/1	
521_0001_FC:2:42:17410:18249#0/1	
;1521_0001_FC:2:35:14465:2698#0/1	
;1521_0001_FC:2:65:15618:6431#0/1	
🧮 🆽 🛄 🕑	

Missed SNP?

chrGCAGAGGCCCAAGCCAGAGGTTCCAGGCTTAAACCCCCAGCCCTGCCCCAGTCCA

ConsensuGCAGAGGCCAAGCCAGAGGTTCCAGGCTTAAACCCCAGCCCTGCCCTGCCCAGTCCA

3866:4795#0/GCA

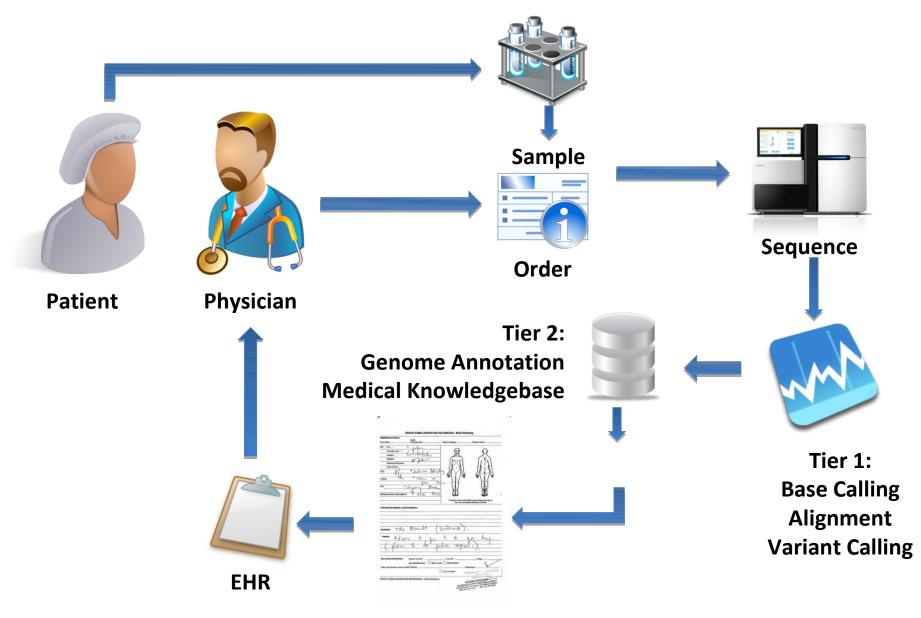
1407:2153#0/

1308:3912#0/GCAGAGGCCAAGCCAGAGGTTCCAGGCTTAAA

Bioinformatics and Health Informatics

If bioinformatics is the study of the flow of information in biological sciences, Health Informatics is the study of the information in patient care

Medicine: Informatics pipeline workflow



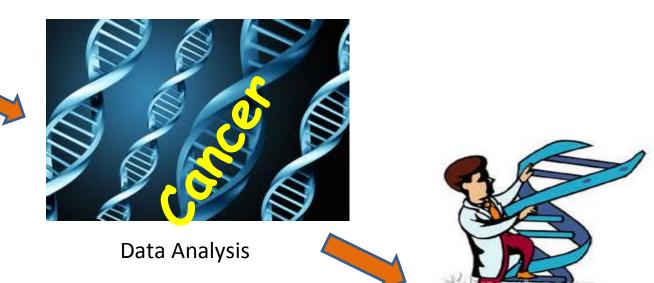
Tier 3: Clinical Report

Huge need in bioinformatics tools

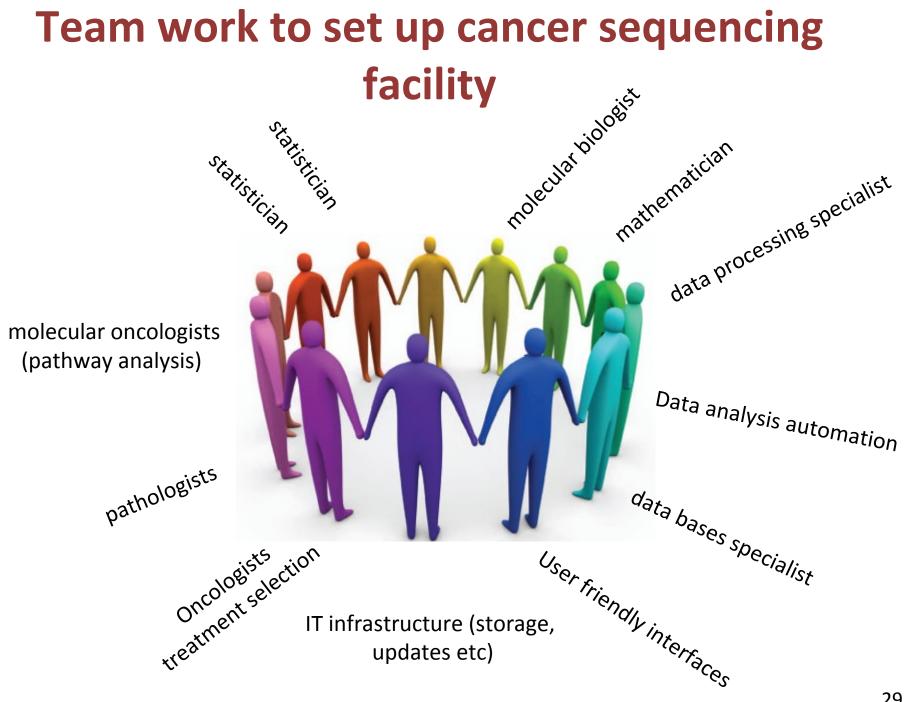
Simple pipelines/protocols and easy to read reports

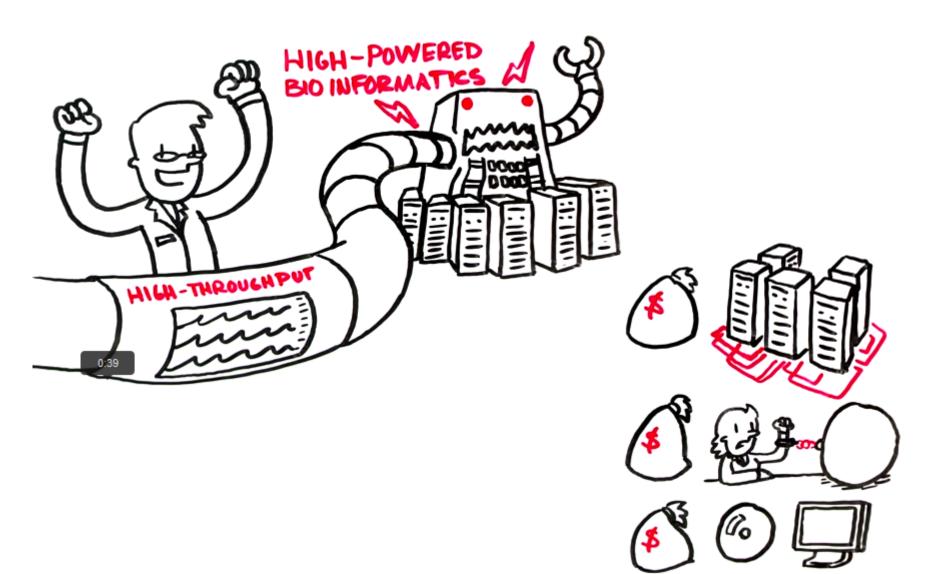


Sample sequencing



Patients treatment





http://www.youtube.com

Ion Torrent: Torrent Suite Software



er Frieder Import	The second secon	Report	Archive
Load data and initiate analysis	Find variants across 1, 2, or 3 samples	Translate variants into digestible reports	Securely manage data, analyses, and reports



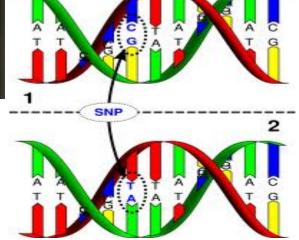
HOMEZ_22814666

	22,814,640 22,814,660 I	22,81. I		DIP	chr14 3TCTTGTATGATCACATCATCATCATCATCA	ATCATCATC	CCTCCTCCTCCTCCTCCTCATCATCTTCTGG
					0/ I#U/2 ATGAGATGATGATGATGATGATGATGA		
		DIP chr14	FCTTGTATGATCACATCATCATCATCATCATCATCATC	TCCTCCTCCTCCTCCTCCTCCTCATCAT		ATCATCATCTTCCTC	CCTCCTCCTCCTCCTCCTCATCATCTTCTGG
				i i	724#0/2		
hr14 GTC	CTTGTATGATCACATCATCATCATCATCATCATCAT	rcifcctcctcctcctc		DIP			CCTCCTCCTCCTCTTCCTCATCATCTTCTGG(
		, DIP				-TCATCATCATCTTC	CCTCCTCCTCCTCTTCCTCATCATCTTCTGGC
		ensus	JTCTTGTATGATCACATCATCATCATCATCATCATC	T <mark>T CCTCCTCCTCCTCCTCCTCCTCATCAT</mark>			CCTCCTCCTCCTCCTCCTCATCATCTTCTGG
	CTTGTATGATCACATCATCATCATCATCATCATCAT			CCTCATCAT	1572#0/1 JTCTTTTTTTTA		-CTCCTCCTCCTCCTCTTCCTCATCATCTTCTGGC
		5#0/1			178#0/1 TCTTCCTTGATTAC		
	-TTGTATGATCACATCATCATCATCATCATCATCATCA	3#0/1	TCTTGTATGATCACATCATCATCATCATCATCAT	T TCCTCCTCCTCCTCCTCCTCATCAT	1970#0/2 JTCTTGTATGATCACA		
	CTTGTATGATCACATCATCATCATCATCATCATCATCA CTTGTATGATCACATCATCATCATCATCATCATCA	7#0/1	FCTTGTATGATCACATCATCATCATCATCATCATCAT			ATCATCATCTTCCTC	CCTCCTCCTCCTCCTCCTCATCATCTTCTGGC
	CCTGTATGATCACATCATCATCATCATCATCATCATCATCATCATCAT	2#0/1	3TCTTGTATGATCACATCATCATCATCATCATCAT	тсстсстсстсстстс	592#0/2 JTCTTGTATGATCACATCATCATCATCAT		CTTCCTCATCATCTTCTGGC
	CTTGTATGATCACATCATCATGCT	0#0/1		3	481#0/1 JTCTTGTATGATCACATCATCATCATCATCA	ATCATCATCTTCCTC	00
	CTTCTATGATCACATCATCATCATCATCATCATCAT	2#0/2	ATC	ATCTTCCTCCTCCTCCTCCTCCTCATCAT	500#0/2 }TCTTGTATGATCACATCATCATCATCATCA	ATCATCATC	CTCCTCCTC
	CTTGTATGATCACATCATCATCATCATCATCATCATCATCA	и и и и и и и и и и и и и и и и и и и	GTATGATCACATCATCATCATCATCATCATCATC	TTCCTCCTCCTCCTCTTCCTCATCAT			CTCCTCCTCTTCCTCATCATCTTCTGGC
	CTTGTATGATCACATCATCATCATCATCATCATCA			AT	068#0/2 }TCTTGTATGATCACATCATCATCATCATCA	ATCATCATC	SCTCCTCCTC
2#0/2	TGTATGATCACATCATCATCATCATCATCATCATCATCA		JICTIGTATCACATCATCATCATCATCATCA				CTCCTCCTCTTCCTCATCATCTTCTGGC
2#0/2	TGTATGATCACATCATCATCATCATCATCATCATCAT		3TCTTGTATGATCATATCATCATCCTCATCATAAACATC	TCCTTCTCCTC	1038#0/2 STCTTGTATGATCACATCATCATCATCATCA	ATCATCATCTTCCTC	
5#0/1	TATGATCACATCATCATCATCATCATCATCAT	гст <mark></mark> тсстсстсст с ^{3#0/2}	CTTGTATGATCACATCATCATCATCATCATCATCATCATCATC	T TCCTCCTCCTCCTCCTCTTCCTCATCAT			CTCCTCCTCTTCCTCATCATCTTCTGG
2#0/1		тсстсст(^{6#0/1}	GTATGATCACATCATCATCATCATCATCATCATCATC	T TCCTCCTCCTCCTCCTCTTCCTCATCAT'	411#0/13101101A10A10A00A10A10A10A10A10A10A	ATCATCATCTTCCTC	
3#0/1		тсстсст (1#0/1	GTATGATCACATCATCATCATCATCATCATCTTC	TTCCTCCTCCTCCTCCTCCTCATCTTCTT			CCTCCTCCTCCTCCTCTTCCTCATCATCTTCTGGC CCTCCTCCTCCTCCTCTTCCTCATCATCTTCTGGC
7#0/1		тсстсст (2#0/2					CCTCCTCCTCCTCCTCTTCCTCATCATCTTCTGGC
	5						
• •			,				
	<u> </u>			c /		c2	
1	S2			s4		s3	
	S2 S8			s4 s5		s5 s7	
		NP	2				
	S8	NP	t	s5		s7	
	S8	NF TCCTCCTCCTCCTCCT(chr14	STCTTGTATGATCACATCATCATCATCATCATCATCATCATCATCATCAT	S5		s7	TCCTCCTCCTCCTCCTCCTCCTCATCATCTT
50#0/1	S8 GTATGATCACATCATCATCATCATCATCATCATC			S5 NP TCCTCCTCCTCCTCCTCCTCCTCATCAT	16#0/1	S7	
50#0/1 85#0/1	S8 GTATGATCACATCATCATCATCATCATCATCATC	TCTTCCTCCTCCT6#0/2		S5 NP TCCTCCTCCTCCTCCTCCTCCTCATCAT	16#0/1	S7	
50#0/1 85#0/1 91#0/2	S8 GTATGATCACATCATCATCATCATCATCATCATC	тс <mark>т</mark> тсстсстсстсст ^{46#0/2} '2#0/1		S5 NP TCCTCCTCCTCCTCCTCCTCCTCATCAT(3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3	16#0/1 —	S7	TCCTCCTCCTCCTCCTCCTCCTCATCATCTT TCCTCCTCCTCCTCCTCCTCCTCATCATCTT TCCTCCTCCTCCTCCTCCTCCTCATCATCTT
35#0/1 35#0/1 35#0/2 27#0/2	S8 GTATGATCACATCATCATCATCATCATCATCATC	TCTCCTCCTCCT46#0/2 '2#0/1 14#0/1		S5	16#U/1 —АТСАТ/ /7#0/2 /СФСАФТСТТ —АТСАТ/ /3#0/2 .СТСАБТСТТ —АТСАТ/ /2#0/2 /СТСФБТСТТБТАТБАТСАСАТСАТСФТСАТ/	S7	
50#0/1	S8 gtatgatcacatcatcatcatcatcatcatcatcatcatcatcat	тс <mark>т</mark> тсстсстсстсст ^{46#0/2} '2#0/1		S5	16#U/1 —ATCATI 17#0/2 /C©CA©TCTT —ATCATI 13#0/2 /CTCAGTCTT —ATCATI 12#0/2 /CTC©GTCTTGTATGATCACATCATC©TCATI 19#0/2 /CTC©GTCTTGTATGATCACATCATC©TCATI	S7	
50#0/1	S8 GTATGATCACATCATCATCATCATCATCATCATC	TCTCCTCCTCCT640/2 2#0/1 4#0/1 CTCCT(1#0/1 2#0/2	TCATCATCATCATCATCATCATCATCATCATCATCATCAT	S5	16#U1	S7	T COTOCT COT COT COT COT COT COT CAT CAT CAT T COT COT COT COT COT COT COT COT COT CAT CAT CAT COT COT COT COT COT COT COT COT COT COT
35#0/1 II 35#0/1	S8 GTATGATCACATCATCATCATCATCATCATCATCATCA ATCATCA GTATGATCACATCATCATCATC GTATGCTCACATCATCATCATCATCATCATCATCATCATCATCAT	тстстсстсстсби0/2 2#0/1 ін#0/1 Стсст(1#0/1 :0#0/2 і9#0/2		S5	1989/01 ATCATI 1780/2 CCCACTCTT ATCATI 1280/2 CTCAGTCTTGTATGATCACATCATCCTCATI 1980/2 CTCAGTCTTGTATGATCACATCATCATCATCATCATCATICATCATICATCATCATCATCATCATCATCATCATCATCATCATCATC	S7	TCCTCCTCCTCCTCCTCCTCATCATCTT TCCTCCTCCTCCTCCTCCTCCTCATCATCTT ATCTT
35#0/1 II 35#0/1	S8 GTATGATCACATCATCATCATCATCATCATCATCA ATCATCA GTATGATCACATCATCATC	тстстсстсстсби0/2 2#0/1 ін#0/1 Стсст(1#0/1 :0#0/2 і9#0/2		S5	1989/01 ATCATI 1780/2 CCCAGTCTT ATCATI 1280/2 CTCAGTCTTGTATGATCACATCATCCTCATI 1980/2 CTCAGTCTTGTATGATCACATCATCATCATCATCATICATICATICATICATIC	S7	TCCTCCTCCTCCTCCTCCTCCTCATCATCTT TCCTCCTCCTCCTCCTCCTCCTCCTCATCATCTT
00#U/1 35#0/1	S8 GTATGATCACATCATCATCATCATCATCATCATCATCATCA GTATGATCACATCATCATCATC GTATGCTCACATCATCATCATCATCATCATCATCATCATCATCAT	ТСТССТССТССТССТК6#0/2 2#0/1 !4#0/1 СТССТ(1#0/1 !0#0/2 !9#0/2 ТСТССТССТССТССТСК6#0/2		S5 NP TCCTCCTCCTCCTCCTCCTCCTCATCAT	1989/01 ATCAT 1780/2 CCCAGTCTT ATCAT 1980/2 CTCAGTCTGTATGATCACATCATCCTCAT 1980/2 CTCAGTCTTGTATGATCACATCATCATCATCATCAT 1980/2 CTCAGTCTTGTATGATCACATCATCATCATCATCAT 1980/2 CTCAGTCTTGTATGATCACCATCATCATCATCAT 1980/2 CTCAGTCTTGTATGATCACATCATCATCATCATCAT 1980/2 CTCAGTCTTGTATGATCACCATCATCATCATCATCATCATCATCATCATCATCA	S7	TCCTCCTCCTCCTCCTCCTCCTCATCATCTT TCCTCCTCCTCCTCCTCCTCCTCCTCATCATCTT ATCTT ATCTT TCCTCCTCCTCCTCCTC
35#0/1	S8 GTATGATCACATCATCATCATCATCATCATCATCATCATCA GTATGATCACATCATCATCATC GTATGCTCACATCATCATCATCATCATCATCATCATCATCATCAT	ТСТ ТССТССТССТССТ(6#0/2 2#0/1 /##0/1 //#0/1 //#0/2 //#0/2 //#0/2 //#0/2 //#0/2 //#0/2 //#0/2 //#0/2 //#0/2 //#0/2 //#0/2 //#0/2 //#0/1 //#0/1 //#0/2 //#0/1 //#0/2 //#0/1 //#0/2 // //#0/2 // //#0/2 // //#0/2 // //#0/2 // //#0/2 // //#0/2 // // //#0/2 // // // // // // // // // // // // /		S5 NP TCCTCCTCCTCCTCCTCCTCCTCCTCATCAT TCCTCCTCCTCCTCCTCCTCCTCCTCATCAT TCCTCCTCCTCCTCCTCCTCCTCCTCATCATA TCTTCCTCCTCCTCCTCCTCCTCCTCCTCCTCCTCCTCC	198901	S7	TCCTCCTCCTCCTCCTCCTCCTCCTCCTCCTCTTC TCCTCC
35#0/1 35#0/1 35#0/1 35#0/1 30#0/2 27#0/2 27#0/2 26#0/2 TTG 26#0/2 TTG 26#0/2 TTG 26#0/2 TTG 28#0/2 TTG 28#0/2 TTG	S8 GTATGATCACATCATCATCATCATCATCATCATCATCATCA GTATGATCACATCATCATCATC GTATGATCACATCATCATCATCATCATCATCATCATCATCATCAT	TCTTCCTCCTCCTCCT680/2 290/1 440/1 		S5 NP TCCTCCTCCTCCTCCTCCTCCTCCTCATCAT TCCTCCTCCTCCTCCTCCTCCTCCTCCTCATCAT TCTTCCTCCTCCTCCTCCTCCTCCTCCTCCTCATCAT TCTTCCTCCTCCTCCTCCTCCTCCTCCTCCTCCTCATCAT	188901 ATCATI 77802 CECAETCTT ATCATI 38902 CTCAGTCTTGTATGATCACATCATCETCAT 28902 CTCEGTCTTGTATGATCACATCATCETCAT 198002 CTCAGTCTTGTATGATCACATCATCATCATCATCATCATCATCATCATCATCAT	S7	TCCTCCTCCTCCTCCTCCTCCTCCTCCTCCTCCTCTTC TCCTCC
U#U/1 1/#U/2 1/#	S8 GTATGATCACATCATCATCATCATCATCATCATCATCATCATCAT	TCTTCTTCTTCTTCT 240/1 440/1 		S5 TCCTCCTCCTCCTCCTCCTCCTCCTCATCAT	18901 ATCAT 77802 CECAETCTT ATCAT 18902 CTCAGTCTTGTATGATCACATCATCETCAT 9802 CTCAGTCTTGTATGATCACATCATCATCATCATCAT 19902 CTCAGTCTTGTATGATCACATCATCATCATCATCATCATCAT 19902 CTCAGTCTTGTATGATCACACATCATCATCATCATCATCATCATCATCATCATC	S7	TCCTCCTCCTCCTCCTCCTCCTCATCATCTT TCCTCCTCCTCCTCCTCCTCCTCATCATCTT ATCTT TCCTCCTCCTCCTCCTCCTCCTCCTCCTCCTC TCCTCCTCCTCCTCCTCCTCCTCCTCCTCCTCCTCCTCC
0000/11 1 11#0/2 12#0/2 12#0/2 15#0/1 16#0/2 TTG 18#0/2 TTG 19#0/2 TTG 19#0/1 TTG 15#0/2 TTG 15#0/2 TTG	S8 GTATGATCACATCATCATCATCATCATCATCATCATCATCATCAT	TCTCTCCTCCTCTCTC TCTCTCTCTCTCTCTCTC 	TCATCATCATCATCATCATC TCATCATCATCATCATCATCATCATCATCATCATCATCAT	S5 TCCTCCTCCTCCTCCTCCTCCTCATCAT	188001 ATCAT 77802 CECAETCTT ATCAT 188002 CTCAGTCTTGTATGATCACATCATCETCAT 189002 CTCAGTCTTGTATGATCACATCATCETCAT 98902 CTCAGTCTTGTATGATCACATCATCATCATCATCAT 198001 CTCAGTCTTGTATGATCACATCATCATCATCATCAT 198001 CTCAGTCTTGTATGATCACACATCATCATCATCAT 198001 CTCAGTCTTGTATGATCACACATCATCATCATCATCATCATCATCATCATCATC	S7	TCCTCCTCCTCCTCCTCCTCCTCCTCATCATCTT TCCTCCTCCTCCTCCTCCTCCTCCTCATCATCTT ATCT ATCT TCCTCCTCCTCCTCCTCCTCCTCCTCCTC TCCTCCTCCTCCTCCTCCTCCTCCTCCTCCTCCTCCTCC
0000/1 1 1 11#0/2 17#0/2 15#0/1	S8 GTATGATCACATCATCATCATCATCATCATCATCATCATCATCA GTATGATCACATCATCATCATC GTATGGTCACATCATCATCATCATCATCATCATCATCATCATCATC	TCTTCTTCTCCTCTC TCTTCTCCTCCTCT 220/1 440/1 90/02 TCTTCTCT(140/1 940/2 1940/2 TCTTCCTCCTCCTCCTS TCTTCCTCCTCCTS TCTTCCTCCTCCTS TCTCCTCCTCCTS TCTCCTCCTCTS 740/2 TCCTCCTCCTCCTS 740/2 TCCTCCTCCTCCTS 740/2 TCCTCCTCCTCCTS 740/2 TCCTCCTCCTCCTS 740/2 TCCTCCTCCTCCTS 740/2 TCCTCCTCCTS 740/2 TCCTCCTCCTS 740/2 750/2 750/	TCATCATCATCATCATCATC TCATCATCATCATCATCATCATCATCATCATCATCATCAT	S5 NP TCCTCCTCCTCCTCCTCCTCCTCCTCATCAT TCCTCCTCCTCCTCCTCCTCCTCCTCATCATAT TCTTCCTCCTCCTCCTCCTCCTCCTCATCATAT TCTTCCTCCTCCTCCTCCTCCTCCTCCTCCTCCTCATCAT	188001 ATCAT 77802 CECAETCTT ATCAT 188002 CTCAGTCTTGTATGATCACATCATCETCAT 189002 CTCAGTCTTGTATGATCACATCATCETCAT 98902 CTCAGTCTTGTATGATCACATCATCATCATCATCAT 198001 CTCAGTCTTGTATGATCACATCATCATCATCATCAT 198001 CTCAGTCTTGTATGATCACACATCATCATCATCAT 198001 CTCAGTCTTGTATGATCACACATCATCATCATCATCATCATCATCATCATCATC	S7	TCCTCCTCCTCCTCCTCCTCCTCCTCATCATCTT TCCTCCTCCTCCTCCTCCTCCTCCTCATCATCTT ATCTT ATCTT TCCTCCTCCTCCTCCTCCTCCTCCTC TCCTCCTCCTCCTCCTCCTCCTCCTCCTCCTCCTCCTCC
SUFU/1 I 35#0/1	S8 GTATGATCACATCATCATCATCATCATCATCATCATCATCATCAT	TCTTCTTCTCCTCTCT TCTTCTCCTCCTCTCT 220/1 440/1 	TCATCATCATCATCATCATCATCATCATCATCATCATCAT	S5 TCCTCCTCCTCCTCCTCCTCCTCCTCATCAT TCCTCCTCCTCCTCCTCCTCCTCCTCATCAT TCCTCCTCCTCCTCCTCCTCCTCCTCCTCATCAT TCTTCCTCCTCCTCCTCCTCCTCCTCCTCCTCCTCCTCC	188011	S7	TCCTCCTCCTCCTCCTCCTCCTCCTCATCATCT TCCTCCTCCTCCTCCTCCTCCTCCTCATCATCT ATCT ATCT CCTCCTCCTCCTCCTCCTCCTCCTCCTCCTCCTCCTCCT
SU#U/1 I 35#0/1	S8 GTATGATCACATCATCATCATCATCATCATCATCATCATCATCAT	TC TCCTCCTCCTCCT 2240/1 ''240/1 ''240/1 ''240/1 ''240/1 ''240/1 ''240/1 ''240/1 ''240/1 ''240/1 ''240/1 ''240/1 ''240/1 ''140/1 ''140/1 ''150/1 ''140/1 ''160/1 ''140/1 ''170/2 ''170/2 TCCTCCTCCTCCTCCT ''140/2 TCCTCCTCCTCCTCCTCTCT ''440/1 ''170/2 ''170/2 TCCTCCTCCTCCTCCTCTCT ''440/1 ''170/2 ''170/2 ''170/2 ''170/2 ''170/2 ''170/2 ''170/2 ''170/2 ''170/2 ''170/2 ''170/2 ''170/2 ''170/2 ''170/2 ''170/2 ''170/2 ''170/2 ''170/2 ''170/2 ''170/2 ''170/2 ''170/2 ''170/2 ''170/2 ''170/2 ''170/2 ''170/2 ''170/2	TCATCATCATCATCATCATCATCATCATCATCATCATCAT	S5 TCCTCCTCCTCCTCCTCCTCCTCCTCATCAT TCCTCCTCCTCCTCCTCCTCCTCCTCATCAT TCTTCCTCCTCCTCCTCCTCCTCCTCATCAT TCTTCCTCCTCCTCCTCCTCCTCCTCCTCATCAT TCTTCCTCCTCCTCCTCCTCCTCCTCCTCCTCCTCCTCC	188011	S7	TCCTCCTCCTCCTCCTCCTCCTCCTCATCATCT TCCTCCTCCTCCTCCTCCTCCTCCTCCTCCTCCTCCTCC
SUFU/1 I 3550/1	S8 GTATGATCACATCATCATCATCATCATCATCATCATCATCATCAT	TC T TCC T CC T CC T (6H0/2 '2#0/1 '4#0/1 '(140/1)	TCATCATCATCATCATCATC TCATCATCATCATCATCATCATCATCATCATCATCATCAT	S5 TCCTCCTCCTCCTCCTCCTCCTCCTCATCAT	INFULT ATCAT 77802 CECAETCTT ATCAT 37802 CTCAGTCTT ATCAT 37802 CTCAGTCTT ATCAT 37802 CTCAGTCTTGATGATCACACATCATCETCAT Intervention 38902 CTCAGTCTTGTATGATCACACATCATCATCATCATCATCATCATCATCATCATC	S7	TCCTCCTCCTCCTCCTCCTCCTCCTCCTCCTCCTCCTCC
U#U/1 15#0/1	S8 GTATGATCACATCATCATCATCATCATCATCATCATCATCATCAT	TC T TCC T CC T CC T CC T (6H0/2 '2#0/1 '4#0/1 'C T CC T (4H0/1 '9H0/2 '9H0/2 '9H0/2 '10T T CC T CC T CC T CC T (4H0/1 '9H0/2 '9H0/2 '9H0/2 '9H0/2 '9H0/2 '9H0/2 '9H0/2 TC T T CC T CC T CC T CC T (4H0/2 T C T CC T CC T CC T (7H0/2 T C T CC T CC T CC T (5H0/2 T C T CC T CC T CC T (5H0/2 T C T CC T CC T CC T (2H0/1) T C T CC T CC T CC T (2H0/1) T C T CC T CC T CC T (2H0/1) T C T CC T CC T CC T (2H0/1) T C T CC T CC T CC T (2H0/1) T C T CC T CC T CC T (2H0/1) T C T CC T CC T CC T (2H0/1) T C T CC T CC T CC T (2H0/1) T C T CC T CC T CC T (2H0/1)	TCATCATCATCATCATCATC TCATCATCATCATCATCATCATCATCATCATCATCATCAT	S5 TCCTCCTCCTCCTCCTCCTCCTCCTCATCAT TCCTCCTCCTCCTCCTCCTCCTCCTCATCAT TCTTCCTCCTCCTCCTCCTCCTCCTCATCAT TCTTCCTCCTCCTCCTCCTCCTCCTCCTCATCAT TCTTCCTCCTCCTCCTCCTCCTCCTCCTCCTCCTCCTCC	INFULT ATCAT 77802 CECAETCTT ATCAT 37802 CTCAGTCTT ATCAT 37802 CTCAGTCTT ATCAT 37802 CTCAGTCTTGATGATCACACATCATCETCAT Intervention 38902 CTCAGTCTTGTATGATCACACATCATCATCATCATCATCATCATCATCATCATC	S7	TCCTCCTCCTCCTCCTCCTCCTCCTCCTCCTCCTCCTCC
U#U/1 15#0/1	SB GTATGATCACATCATCATCATCATCATCATCATCATCATCATCAT	TC T T CC T CC T CC T CC T (6H0/2 '2#0/1 \u00e4400/1 C T CC T (1H0/1 \u00e400/1 '9H0/2 '9H0/2 '9H0/2 '9H0/2 '10T T CC T CC T CC T CC T (2H0/2 '9H0/2 TC T T CC T CC T CC T CC T (2H0/2 CT T CC T CC T CC T CC T (2H0/2 TC T CC T CC T CC T CC T (2H0/2 TC CT CC T CC T CC T (2H0/2 T C CT CC T CC T CC T (2H0/2 T C CT CC T CC T (2H0/2 T C CT CC T CC T CC T (2H0/2 T C T CC T CC T CC T (2H0/2 T C T CC T CC T CC T (2H0/2 T C T CC T CC T CC T (2H0/2 T C T CC T CC T CC T (2H0/2 T C T CC T CC T CC T (2H0/2 T C T CC T CC T CC T (2H0/2 T C T CC T CC T CC T CC T (2H0/2 T C T CC T CC T CC T CC T (2H0/2 T C T CC T CC T CC T CC T (2H0/2 T C T CC T CC T CC T (2H0/2 T C T CC T CC T CC T CC T (2H0/2 T C T CC T CC T CC T CC T (2H0/2 T T C T CC T CC T CC T CC T (2H0/2	TCATCATCATCATCATCATC TCATCATCATCATCATCATCATCATCATCATCATCATCAT	S5 TCCTCCTCCTCCTCCTCCTCCTCCTCATCAT	INFULT ATCAT 77802 CECAETCTT ATCAT 37802 CTCAGTCTT ATCAT 37802 CTCAGTCTT ATCAT 37802 CTCAGTCTTGATGATCACACATCATCETCAT Intervention 38902 CTCAGTCTTGTATGATCACACATCATCATCATCATCATCATCATCATCATCATC	S7	TCCTCCTCCTCCTCCTCCTCCTCCTCCTCCTCCTCCTCC
SU#U/1 I 35#0/1	SB GTATGATCACATCATCATCATCATCATCATCATCATCATCATCAT	TC T TCC T CC T CC T CC T (6H0/2 '2#0/1 '4#0/1 'C T CC T (4H0/1 '9H0/2 '9H0/2 '9H0/2 '10T T CC T CC T CC T CC T (4H0/1 '9H0/2 '9H0/2 '9H0/2 '9H0/2 '9H0/2 '9H0/2 '9H0/2 TC T T CC T CC T CC T CC T (4H0/2 T C T CC T CC T CC T (7H0/2 T C T CC T CC T CC T (5H0/2 T C T CC T CC T CC T (5H0/2 T C T CC T CC T CC T (2H0/1) T C T CC T CC T CC T (2H0/1) T C T CC T CC T CC T (2H0/1) T C T CC T CC T CC T (2H0/1) T C T CC T CC T CC T (2H0/1) T C T CC T CC T CC T (2H0/1) T C T CC T CC T CC T (2H0/1) T C T CC T CC T CC T (2H0/1) T C T CC T CC T CC T (2H0/1)	TCATCATCATCATCATCATC TCATCATCATCATCATCATCATCATCATCATCATCATCAT	S5 NP TCCTCCTCCTCCTCCTCCTCCTCCTCATCAT TCCTCCTCCTCCTCCTCCTCCTCCTCATCATAT TCTTCCTCCTCCTCCTCCTCCTCCTCATCATAT TCTTCCTCCTCCTCCTCCTCCTCCTCCTCCTCATCATA TCTTCCTCCTCCTCCTCCTCCTCCTCCTCCTCCTCCTCC	INSTRUCT ATCATE 77802 CCCAGTCTT ATCATE 37802 CTCAGTCTTGTATGATCACATCATCGTCAT 37802 CTCAGTCTTGTATGATCACATCATCGTCAT 39802 CTCGGTCTTGTATGATCACATCATCATCGTCAT 39802 CTCAGTCTTGTATGATCACATCATCATCATCATCATCATCATCATCATCATCAT	S7	TCCTCCTCCTCCTCCTCCTCCTCCTCCTCCTCCTCCTCC
30#0/1 1 35#0/1	SB GTATGATCACATCATCATCATCATCATCATCATCATCATCATCAT	TC T T CC T CC T CC T CC T (6H0/2 '2#0/1 \u00e4400/1 C T CC T (1H0/1 \u00e400/1 '9H0/2 '9H0/2 '9H0/2 '9H0/2 '10T T CC T CC T CC T CC T (2H0/2 '9H0/2 TC T T CC T CC T CC T CC T (2H0/2 CT T CC T CC T CC T CC T (2H0/2 TC T CC T CC T CC T CC T (2H0/2 TC CT CC T CC T CC T (2H0/2 T C CT CC T CC T CC T (2H0/2 T C CT CC T CC T (2H0/2 T C CT CC T CC T CC T (2H0/2 T C T CC T CC T CC T (2H0/2 T C T CC T CC T CC T (2H0/2 T C T CC T CC T CC T (2H0/2 T C T CC T CC T CC T (2H0/2 T C T CC T CC T CC T (2H0/2 T C T CC T CC T CC T (2H0/2 T C T CC T CC T CC T CC T (2H0/2 T C T CC T CC T CC T CC T (2H0/2 T C T CC T CC T CC T CC T (2H0/2 T C T CC T CC T CC T (2H0/2 T C T CC T CC T CC T CC T (2H0/2 T C T CC T CC T CC T CC T (2H0/2 T T C T CC T CC T CC T CC T (2H0/2	TCATCATCATCATCATCATC TCATCATCATCATCATCATCATCATCATCATCATCATCAT	S5 NP T CCTCCTCCTCCTCCTCCTCCTCCTCATCAT T CCTCCTCCTCCTCCTCCTCCTCCTCATCAT T CTCCTCCTCCTCCTCCTCCTCCTCATCAT T CTTCCTCCTCCTCCTCCTCCTCCTCCTCATCAT T CTTCCTCCTCCTCCTCCTCCTCCTCCTCATCAT T CCTCCC T CCTCCTCCTCCTCCTCCTCCTCCTCCTCCTCATCAT T CCTCCTCCTCCTCCTCCTCCTCCTCCTCATCAT T CCTCCTCCTCCTCCTCCTCCTCCTCCTCATCAT T CCTCCTCCTCCTCCTCCTCCTCCTCCTCCTCATCAT T CCTCCTCCTCCTCCTCCTCCTCCTCCTCATCAT	INFUDI	S7	TCCTCCTCCTCCTCCTCCTCCTCCTCATCATCTT TCCTCCTCCTCCTCCTCCTCCTCCTCATCATCTT ATCT ATCT TCCTCCTCCTCCTCCTCCTCCTCCTCCTC TCCTCCTCCTCCTCCTCCTCCTCCTCCTCCTCCTCCTCC
U#U/1 1 5#0/1	SB GTATGATCACATCATCATCATCATCATCATCATCATCATCATCAT	TC TCCTCCTCCTCCT '2#0/1 '2#0/1 '4#0/1 '19#0/2 '19#0/2 '19#0/2 '19#0/2 TC TCCTCCTCCTCCTCS '9#0/2 TC TCCTCCTCCTCCTCS '9#0/2 TC TCCTCCTCCTCCTCS '9#0/2 TC TCCTCCTCCTCCTCS '#0/2 TCCTCCTCCTCCTCCTS '#0/2 TCCTCCTCCTCCTCTS '#0/2 TCCTCCTCCTCCTCTS '#0/2 TCCTCCTCCTCCTCTS '#0/1 TCCTCCTCCTCCTCTS '#0/1 TCTTCCTCCTCCTCTCT '#0/1 TCTTCCTCCTCCTCTCTCTS '#0/1	TCATCATCATCATCATCATC TCATCATCATCATCATCATCATCATCATCATCATCATCAT	S5 TCCTCCTCCTCCTCCTCCTCCTCCTCATCAT TCCTCCTCCTCCTCCTCCTCCTCCTCATCATAT TCTTCCTCCTCCTCCTCCTCCTCCTCCTCATCATA TCTTCCTCCTCCTCCTCCTCCTCCTCCTCCTCCTCCTCC	INFUDI ATCATC 77802 CECAETCTT ATCATC 37802 CTCAGTCTT ATCATC 38902 CTCAGTCTTGTATGATCACATCATCETCAT ATCATC 39902 CTCEGTCTTGTATGATCACATCATCETCAT Information 39902 CTCAGTCTTGTATGATCACATCATCATCATCATCATC Information 39901 CTCAGTCTTGTATGATCACACATCATCATCATCATCATC Information 39902 CTCAGTCTTGTATGATCACACATCATCATCATCATCATCATCATCATCATCATC	S7	TCCTCCTCCTCCTCCTCCTCCTCCTCCTCCTCCTCCTCC

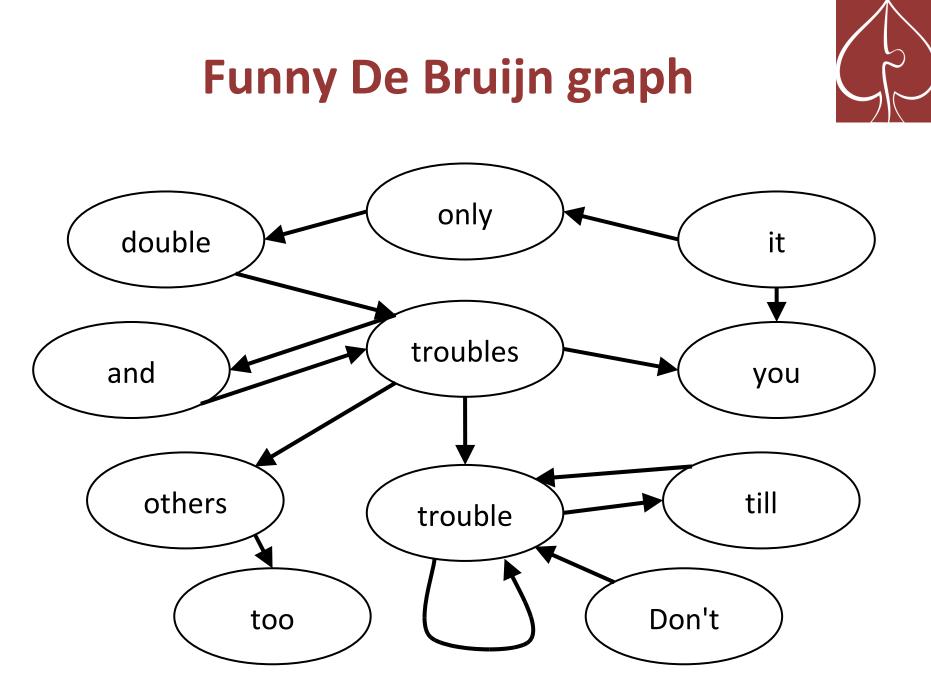
Each baby to be sequenced at birth: personal reference







"GATTACA", 1997



Modified from Andrey Prjibelski





THANK YOU!